Image retrieval through qualitative representations over semantic features

Zia Ul Qayyum

Abstract This paper proposes a qualitative knowledge-driven semantic modelling approach for image understanding and retrieval. The similarity measure is calculated for each query by exploiting the notion of *conceptual neighbourhood*—a measure of closeness between qualitative relations. The relative similarity of two images is proportional to the qualitative similarity measure value. The approach is motivated by the need to bridge the semantic gap between a human user and that of CBIR systems and enable semantic querying in such systems. Local semantic concepts, such as sky, grass, of an image are used to obtain a semantic image description. Four kinds of qualitative spatial representations have been applied to these semantic concepts. This allows for representation and reasoning of an image's content structures at a more abstract level than pixels or other low level features and provides a higher level, semantic basis for image understanding. We investigate whether such a representation of an image's visual content also provides an effective and natural way to provide content-oriented querying. We also investigate whether querying based on multiple representations is effective, and report on three voting schemes used to retrieve images in this way. The test data set having hand-assigned class labels has been used to have a metric evaluation of the retrieval accuracy. The results compare favourably with a non-qualitative representation based on the same semantic features which simply compares the percentage of each feature in pairs of images.

Keywords Image retrieval · Semantic features · Qualitative representations · Qualitative similarity

1 Introduction

Advances in digital technologies along with the growth of World Wide Web have resulted in universal access to very large archives of digital data. This has lead to an increasing requirement for systems which can handle such a dynamic and complex visual content at a higher semantic level. Moreover, more flexible and robust techniques are required in such systems. Content based image classification and retrieval systems, therefore, have gained more importance and have been an active research area in recent years [1]. In all such

Z. U. Qayyum (✉)
Department of Computer Science, National University of Computer & Emerging Science,
Chiniot-Faisalabad Campus, Faisalabad, Pakistan
e-mail: ziaqayyum@gmail.com

systems, image interpretation and understanding plays a vital role. Most of the research in this area is primarily based on use of low level image features like colour, texture, shape etc. [10, 15, 25–27]. Although low level image processing algorithms and methodologies are quite mature, such systems are hard to be used effectively by a novice due to the semantic gap between user perception and understanding, and system requirements. Therefore, bridging this gap between low level synthetic features and high level semantic meanings is generally regarded as an open problem [1]. Humans tend to describe the scenes using natural language semantic keywords/concepts like sky, water etc. and specify retrievals "an image with water next to fields and having sky at the top…." or "… has a small lake with high peaks of mountains behind and fields on left….". This suggests that use of underlying semantic knowledge in a qualitative representation language may provide a way to model the human context and a natural way to bridge this semantic gap for better image understanding, categorization and retrieval capabilities [19].

This paper proposes a qualitative knowledge-driven semantic modelling approach for image categorisation and retrieval. As discussed above, qualitative representation of the local semantic contents of image allows for representation and reasoning of content structures at a higher abstraction level than pixels or other image low level features. In earlier work [19], we showed how category descriptions for a set of images could be learned; using qualitative spatial representations (QSR) over a set of local semantic concepts (LSC) such as sky, grass. There were six global categories (e.g. Coasts, Landscapes with Mountains etc.) [24] and we used three kinds of QSR techniques to demonstrate that supervised learning of a pure qualitative and spatially expressive representation of semantic image concepts can rival a non qualitative approach [19, 31] for image categorization, and moreover result in a more intuitive and more human understandable image description. Details of these qualitative representations are presented later.

In this paper we turn our attention away from categorisation and to the retrieval of images either given example(s), or a qualitative description. We base our work on the same semantic descriptions as in the categorization work summarized above: our hypothesis is that the qualitative representations which were able to effectively support categorization may also provide an effective and natural way to support content-oriented querying approach. A query can either be directly described in the qualitative representation or a query can be given as a sample image (i.e. query by example: QBE)—the system then forms a qualitative description of it and compares this with qualitative descriptions of images in the database of images, and uses a *qualitative similarity measure* to retrieve qualitatively similar images [20]. The qualitative similarity measure is based on the notion of a conceptual neighbourhood [11], discussed in more detail below in Section 4. The relative level of similarity is proportional to value of the similarity measure. Therefore, a sorted list of this measure corresponds to the respective image's level of similarity according to this order. The retrieved images can be grouped together into "Most Typical", "Typical", and "Less Typical" images by using certain thresholds on the similarity measures—these thresholds may be qualitatively determined by inspection of the qualitative behaviour of histograms for buckets of images over the similarity ordering. In order to evaluate the performance of this approach to image retrieval, we take advantage of the manually assigned categories for the image data base in our experiments. Although we are not performing image categorization, and the retrieval algorithm does not use the category information, we can evaluate the success of a retrieval by counting the number of images in the same category as the query image near the top of the rank ordering of retrieved images.

In experiments using this technique on the different qualitative representations we observed that different measures have different levels of performance for different categories

of images; this lead us to also investigate the use of voting schemes in order to combine the different qualitative representations to enhance the performance of the retrieval system overall.

The potential advantages of this approach are threefold. Firstly, a qualitative description is arguably closely resembles human cognition in such domains and, therefore, may also model human perception in this context. Secondly, qualitative and semantic representation of image content provides the opportunity to specify a query either with an example image, or with a qualitative description over semantic concepts. Thirdly, the approach does not rely on existing segmentation techniques to learn semantic labels using low level features of an image; our approach is region-based and is segmentation free.

In the work described here, a collection of 700 natural scenes images has been used. The labelled data set was provided by Julia Vogel who had developed a semantic modelling framework for image categorisation and retrieval [31]. Our approach builds on her work; a brief overview of her approach is presented in Section 2.4.

The rest of the paper has been structured as follows. Related work is briefly discussed in Section 2. Section 3 describes our approach for image description using qualitative representations. A qualitative similarity based image retrieval approach is presented in Section 4. Section 5 presents the results and evaluation of the approach, while Section 6 presents our conclusions and suggestions for future work.

2 Related work

In the image retrieval literature, image description and better understanding of underlying semantic content plays an important role as the nature and structure of the query mainly depends on the underlying image description. Moreover, image categorization may provide a keyword based querying facility using global image labels in content-based image categorization and retrieval systems. The following subsections, therefore, describe some relevant work from allied disciplines of content-based image retrieval.

2.1 Image description/categorization

As discussed above, much previous research on image description and categorization has been accomplished through low-level feature vectors like colour, texture and shape, whereas semantic scene description is arguably a natural way to describe image features and it may bridge the gap between a human's description and that of a computer. In most of the literature, semantics is only found in definitions of scene classes such as indoor vs. outdoor, city vs. landscapes, mountain vs. forest etc. while classification itself is based on low level image features. These approaches are based on the assumption that the images with similar colour and texture features are semantically closer.

Vailaya et al. [30] describe a hierarchical scheme for classifying vacation images using colour and edge direction features. They classify at a first step images into city vs. landscapes; the landscapes category is further classified into sunset, forest and mountain classes. The probabilistic model of low level features required for the Bayesian framework is estimated using vector quantization. Most of the techniques in this category of research use low level image features and some use some spatial content as well. Moreover, in some cases categorization proceeds hierarchically to further subdivide these classes but work on categorization into multiple classes simultaneously is sparse. In these binary approaches, the rate of accuracy is quite good perhaps not surprisingly since it is likely there will be less variation

of low level concepts along only two branches of categorization. Our approach and that of [31], on the other hand, provides a classification scheme for categorizing and retrieval of natural scene images into six semantically meaningful classes based on local semantic image content.

Semantic image description may improve image description and understanding significantly and can bridge the semantic gap between humans and computer systems. As an example of this approach to image categorization, based on local semantic contents of an image, Serrano et al. [24] use two semantic attributes, namely sky and grass, and low level colour and wavelet texture features to classify images into indoor vs. outdoors using support vector machines (SVM) with an overall accuracy of 87.2 %. The semantic scene attributes, sky and grass, were predicted using the same low level features and integrated into the classification scheme already learnt for an improved two step indoor/outdoor classification scheme using a Bayesian network which resulted in an improved classification rate of 90.7 %. This shows that the use of low level semantic concepts in image description may improve expressiveness and classification accuracy, in line with our proposed approach. Maron et al. [16] have described a method for categorizing natural scenes into waterfall, mountain and fields. The images are modelled as bags of multiple instances (sub regions) and a bag is labelled as positive if at least one instance in the bag is positive and negative otherwise. The model learns scene templates for each class and then a probabilistic diversity density method is used to learn concepts from multiple instance examples. The classification results are evaluated based on the RGB colour features of an image that are closer to at least one positive instance in every positive labelled bag of instances for a class. Ciocca et al. [6] have also presented an image retrieval approach using prosemantic features. The objective of this approach was mainly based on incorporating the semantics into the image description and categorization process.

Segmentation of image regions for feature extraction is another important technique, which is widely discussed in the literature and used in image retrieval and computer vision context. Depalov et al. [9] present an approach for segmenting and hierarchically labelling natural scenes images into perceptually and semantically uniform regions by using texture and spatially adaptive colour features. Regions are hierarchically labelled using dominant semantic concepts such as natural, man-made, human, animal etc. with sub-categories like vegetation, sky, building etc. The segmentation algorithms, though are quite mature, still tend to over or under segment the image, constraining the overall accuracy of the system by the accuracy of region segmentation [31]. By contrast, the semantic annotation scheme we use here does not rely on variable segmentation techniques; rather, it is based on a fixed segmentation of the image into a 10×10 grid resulting in 100 patches which are labelled with semantic concepts; these semantic concepts are extracted using low level features as detailed below in Section 2.4.

Image annotation has also been used to obtain a better image description and classification accuracy. Picard et al. [18] introduced the idea of annotating image regions using texture. The framework selects a best model from a multitude of texture models to annotate the image regions with semantically meaningful labels. It first learns from the user's input and interaction and then propagates the learnt labels to other similar regions. Town et al. [29] developed a system to classify segmented image regions into semantic labels using a neural network, segmenting the image using colour and texture features. This approach and the one presented here are similar in using semantic labels but labelling of image regions in our approach can be described as to be region based as opposed to the segmentation approach based on low level colour and texture features.

From the above discussion, it appears that most of the existing techniques in image description and image classification are predominantly quantitative and based on describing and categorizing images using low level features such as colour, texture [17, 25, 27, 30]. Scene descriptions are typically not expressed in terms of underlying semantic knowledge or

using qualitative and spatial relationships between attributes used for image description. Our approach addresses these issues by applying a variety of qualitative spatial representations onto semantically annotated image regions/patches of each of the semantic concepts in an image. This abstracts away from dealing with low level image features and provides higher level, more expressive descriptions, which are arguably more intuitive and spatial, and potentially useful for semantic querying and image retrieval. This representation has been used in different experiments to have a more intuitive and expressive categorization of images into six semantically meaningful classes [19].

2.2 Image retrieval using semantic image description

Content-based image retrieval systems have been an active research area in computer vision during recent years. A more recent review of CBIR techniques is Singhai and Shandilya [26] which highlights the importance of using higher semantic content along with low-level features to model human perception in image retrieval process. Another study is by Deb et al. [8], which discusses the state of the art in segmentation, indexing and retrieval techniques in a number of CBIR systems. They note that despite much work in aspects related to high level semantics of image features, the gap between low level image features and high level semantic expressions are bottlenecks to the access of multimedia data from databases. Madugunki et. al. [15] have also published a paper to present classification of CBIR systems. Using number of low level features like local and global color histograms, HSV, DCT and DWT features to compare the results of CBIR systems. These surveys all reveal one important aspect that almost all existing approaches rely on using low level image features for image description, categorization and retrieval. Since image understanding is a key to all content-based image categorisation and retrieval systems, so a human understand-able image description may yield more robust systems since humans normally tend to use semantic and qualitative terms to describe a situation/image. Therefore, a retrieval system based on underlying semantic knowledge may help a non-expert user to use such systems more effectively. In this category, some research has been done focusing on the use of labelling the image regions with semantic concepts and carrying out key-word based search for image retrieval. Bradshaw [4] proposed a probabilistic approach to assign small image areas labels such as "man-made" and "natural", and global labels such "inside", "outside" etc. to whole images using class likelihoods from colour-texture features of images for a semantic image retrieval. Town et al. [29], as mentioned above, have also have annotated local regions of images with 11 and 10 semantic categories respectively, Town et al. do not assign a global label to the images, so retrieval is based on local semantic concepts only. Aghbari et al. [1] also demonstrate an image retrieval approach based on semantically labelled image regions. These image regions have been hierarchically classified based on their semantics using low level image features. The retrieval is based on these semantic keywords attached to particular images. Enser et al. [10] present a survey of issues relevant to the semantic gap in image retrieval systems. They focused particularly on the nature and representation of semantic content of an image in image retrieval systems.

Wang et al. [32] discuss an approach for semantic retrieval based on content and context of image regions and which supports both keyword and QBE approaches. In this approach, images are segmented using a semantic codebook based on colour and texture classification. The content and context describe a region's low level features and their relationships respectively. It uses only dominant semantic categories of an image and most typical images in that category are selected manually from an image database which can best model the codebook representing colour and texture classification for that particular semantic category.

Another query by semantic example (QBSE) approach is proposed by Rasiwasia et al. [22] based on posterior concept probabilities of each concept in an image. QBSE is accomplished by comparing the probability simplexes of the query image and all database images to find the closest neighbours. The perceptual segmentation approach in Depalov et al. [9], as discussed in the Section 2.1 above, has not been applied in their work for image categorization and retrieval; but the relative effectiveness of their approach in regards to image segmentation and labelling can be used to perform keyword based image retrieval. The VISENGINE system of Sun at al [28] relies on segmenting image regions by clustering visual features like colour, texture, shape etc. and differentiating them into foreground and background regions. A semantic visual template for each of the background and foreground regions is used to retrieve images. User feedback on retrieved images is gathered to mark "most relevant" images with respect to each query image. A visual template of features generated by mining the relationship between feature weights of all retrieved images and those of marked by the user is used to try to improve the similarity level at the next iteration.. The approach is largely user-centered, and therefore results may vary depending on human perception and context. Only large regions are identified during segmentation which inhibits a true semantic similarity in the retrieved images, as relatively small image areas do not contribute towards the retrieval process. CIRES is an online CBIR system which uses image structures in addition to colour and texture features to achieve a more robust image retrieval framework [14]. Perceptual groups of hierarchically extracted image structures like lines, segments, long linear lines, L-junctions, U-junctions, polygons etc. are formed to create high level meaningful concepts. The approach is claimed to be particularly helpful in retrieval of images containing man made objects like windows, walls etc., as they contain such structural objects. Howe [12] has used different algorithms combined together to boost the classification and retrieval accuracy. Sebe et al. [23] have also surveyed recent approaches and the state of the art in the areas of semantic image/video retrieval, interactive retrieval frameworks, retrieval based on human perception, and relevance feedback strategies in information retrieval. The use of ontologies and metadata representation languages is another recent trend for annotating and retrieving images [13]. A prerequisite for applying this approach is the construction of generic and possibly domain specific ontologies and detailed annotations.

One crucial research question for QBE systems is how to measure the level of similarity, and assess the accuracy of such a technique. Defining a notion of similarity is fraught with difficulties as context may play a pivotal role. Moreover, when using a qualitative representation, where feature descriptions do not take quantitative values, the very notion of a metric becomes problematic; approaches to qualitative similarity are discussed in [3]. In computer vision and image processing, metric approaches have generally been used to compute scene similarity. Vogel et al. [31] use a semantic typicality measure based on normalised distance for a semantic ordering of natural scenes in categories such as forest and mountains, mountains and rivers/lakes. Indeed, in most CBIR systems, similarity is based on a metric evaluation and images with desired level of image features, either low level or semantic concepts are retrieved. Burn et al. [5] use the concept of "gradual change" to determine spatial similarity of scenes. This metric is based on the gradual transformation of spatial relations, namely topological, directional and distance relations, in an attempt to transform one scene into the other. The number of transformations required determines the level of spatial similarity between scenes. Each transformation can be thought of as changing one relation to a conceptual neighbour of the relation. The notion of conceptual neighbourhood was first put forward by Freksa [13] in the context of a set of 13 pairwise and disjoint relations between temporal intervals and was defined as "two spatial or temporal relations are conceptual neighbours if one can be transformed into the other by a single transformation/transition" (also related to the continuity networks described in [21]. Our approach to measuring qualitative

similarity is also based on the use of the use of conceptual neighbourhoods as discussed further below in Sections 4 and 5.

2.3 Qualitative spatial representations

There has been an increasing amount of research into qualitative spatial representation and reasoning in the fields of AI and other disciplines including Geographic Information Science as it can be argued to provide an appropriate cognitively or intuitively relevant representation for spatial information—typical spatial expressions in natural language are qualitative rather than quantitative; moreover qualitative representations abstract away from quantitative computation, and from noise and uncertainty in perceptual data. It has widely been used in different application domains like GIS, NLP, robotics, computer vision etc. Cohn and Hazarika [7] review the state of the art in qualitative spatial representation and reasoning and related issues with identification of some application areas as well. There are now many qualitative spatial calculi, covering aspects such as topology, distance, orientation, and shape. Rather than attempt an exhaustive analysis of the utility of all these calculi, we concentrate on a small set of qualitative spatial relations here; we do not claim these are necessarily the best calculi for image description, or even for the particular kinds of images in the database we use here, but we leave that for further work. Our aim is simply to illustrate the use of qualitative calculi for image retrieval and to demonstrate their potential applicability.

Initially we will use the three qualitative representations we used in our earlier work on image categorization using QSR [19]: Allen's interval calculus [2], Chord patterns [17] and relative size. We claim that these representations are sufficient to demonstrate that the use of qualitative relations between semantic regions can provide an effective and natural way to support content-oriented querying approach. Although that Allen's calculus was originally intended for temporal representations, it can also be used for representing 1D space. These three representations are discussed further in Section 3.

2.4 Baseline approach

Our approach builds on Vogel et al's work [31]; a brief description of her work is presented in this section to make this paper self-contained. In their approach, images were divided into a grid of 10×10 regions to extract local image regions. By analyzing these regions in images, nine local[1] and discriminating semantic concepts were identified: sky, water, grass, foliage, flowers, field, mountain, snow, trunks and sand. Using these labels, 99.5 % of the images were manually annotated. Subsequently supervised machine learning techniques were developed to automatically annotate images—we use the original hand labelled data set in our work here in order to concentrate our evaluation on the semantic and qualitative representations. A label "rest" is used for unidentified patches or occurrences of other semantic categories. A sample description is presented in Fig. 1. Based on this representation, the percentages of concept occurrences, concept occurrence vector (COV), were evaluated on between three and five horizontally divided image regions (e.g. top/middle/bottom). Images were represented by frequency histograms of local semantic concepts such as grass, foliage, water etc. and based on a semantic typicality measure, the images were categorized into one of the six semantically meaningful categories sky_clouds, coasts, landscapes_with_mountains (lwm), fields, forests, waterscapes.

---

[1] Vogel has cited nine semantic concepts in her work [31], while the data set she provided and which has been used in experiments contains two extra ones (mountain and snow)—however these occur infrequently and the basis for comparison will be largely unaffected.
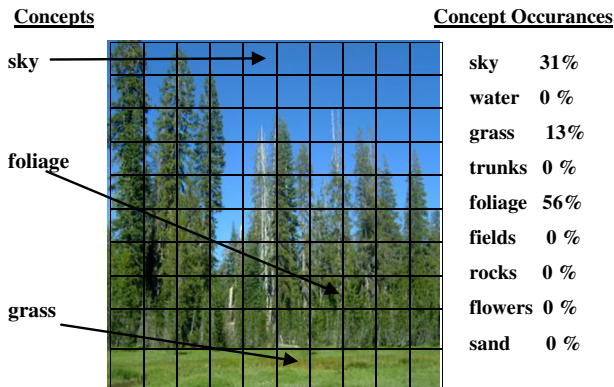
Concepts   **Concept Occurances**

sky	31%
water	0 %
grass	13%
trunks	0 %
foliage	56%
fields	0 %
rocks	0 %
flowers	0 %
sand	0 %

Fig. 1 A segmented image with 9 local semantic concepts and COV

Their approach is partially spatial through its division of the image into horizontal bands (e.g. top/middle bottom) but is mainly based on the metric value of the percentages of discriminant semantic concepts.

Following on from our earlier mentioned above [19], where we used a qualitative representation over these semantic categories for learning image categories, in the work we report here we conduct a variety of experiments to retrieve images labelled with the same semantic concepts but using qualitative spatial similarity measures over these concepts rather than computing percentages. The details of our methodology for image description and retrieval are discussed in the next section. A sample from the image data set used representing each of six classes is displayed in Fig. 2[2],, which illustrates the six classes/categories with which all 700 images have been hand labelled.

3 Qualitative image description

In this section, we present our qualitative approach based on local semantic concepts for image description. The approach is briefly explained here—further details can be found in [19]. We use the same three kinds of qualitative spatial relations already used in our earlier work on categorization:

1) The relative size (*'RSizeRep'*—measured in grid squares) of each of the concept occurrences in each image. The relative size is calculated for all possible pairwise combination of semantic labels. Since there are 11 labels and only one ordering needs to be considered, this gives 66 pairings. Each may be regarded as an attribute of the image with possible values of 'Greater than' (>), 'Less than' (<) and 'Almost Equal to' (≈). These relations are defined as:

    Let |P| denote the number of occurrences of a patch type 'P' (a patch refers to one grid square containing a semantic concept, as represented in Fig. 1 above). If 'P1', 'P2' are two patch types, then:

    •   P1>P2 iff (0.9*|P1|>|P2|)

---

[2] In fact for copyright reasons, we have substituted similar images in this paper, and the percentages in Fig. 1 have been hand estimated.
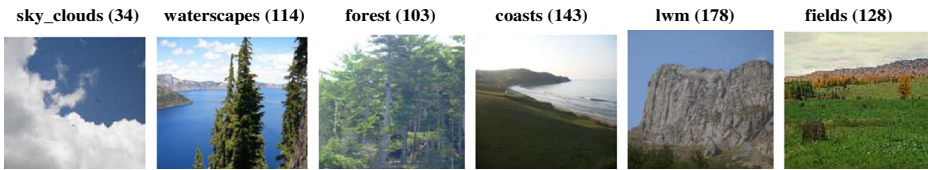
sky_clouds (34) **waterscapes (114)** **forest (103)** **coasts (143)** **lwm (178)** **fields (128)**

Fig. 2 Sample images of the six categories of natural scenes

- P1<P2 iff (1.1* |P1|<|P2|)
- P1≈P2 otherwise
  Note that we have used a tolerance of ±10 % since it is relatively unlikely that two attributes/labels would ever have exactly equal size in similar images.

2) Allen relations [2] (*'AllenRep'*—measured on vertical axis between the intervals representing the maximum vertical extent of each concept occurrence). Allen's calculus has been used to represent 1D spatial knowledge using thirteen relations namely, 'before' (<), 'meets' (m), 'overlaps' (o), 'during' (d), 'starts' (s), 'finishes' (f) and their inverses 'after' (>), 'met-by' (mi), 'overlapped-by' (oi), 'contains' (di), 'started-by' (si), 'finished-by' (fi) respectively, and 'equal' (=). If either or both of the semantic attributes being compared using Allen's relations are missing in the image, we indicate this with "No" in place of the Allen relation. Again this gives 66 pairings describing each image.

3) Chord patterns [17] (*'ChordRep'*) of semantic concepts applied to each grid row. In this approach, each semantic feature is a 'tone' and the set of these in each row forms a 'chord'. Thus, given the 10×10 grid that describes the semantic attributes for each image, we thus generate 10 chords, one for each row to describe each image such as "foliage sky" or "grass sky sand water" etc[3]. This representation thus generates set-valued attributes.

In addition to the above three representations already used in our categorization work [19], we add a fourth representation: whether one patch type is spatially in contact with another in the image (*'TouchRep'*). Note that the Allen meets relation does not guarantee this since the two patches may be at different sides of the picture, nor does the Allen representation explicitly record horizontal contact; similarly the fact that two semantic concepts are present in adjacent chords does not guarantee contact either. To represent genuine touching relationships we thus introduce 55 pairwise comparisons to record which patch types touch which other ones.

In addition to these qualitative representations, for comparison purposes, we also ran experiments with a purely quantitative metric based retrieval scheme based on the respective percentages of each of the semantic concepts in each image has also been investigated in the style of Vogel et al. [31][4].

Figure 3 shows a segmented image described by the relative size and Allen relationships while Fig. 4 illustrates an image described with the fine grained chord representation.

---

[3] This representation can be regarded as an abstraction of the relation used by Vogel et al. above—whereas they record the percentage of each attribute in each horizontal band, in the chord representation it is only presence or absence which is recorded.
[4] This representation is labelled as "%ages" in Tables 2, 3 and 4 below. Similarity is computed using the sum of absolute differences in percentage values for each attribute in a pair of images.
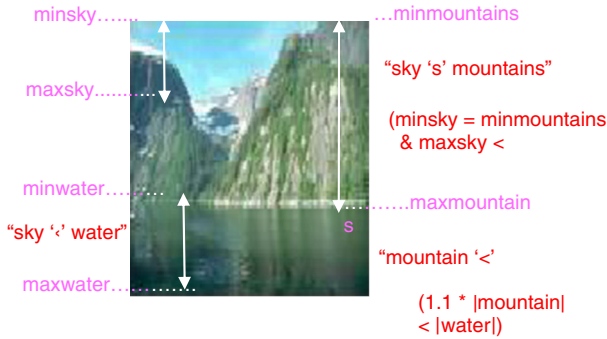
minsky…....   …minmountains

maxsky.......   "sky 's' mountains"

                                               (minsky = minmountains
                                                & maxsky <

minwater…....   …….maxmountain

"sky '<' water"   s

maxwater…...   "mountain '<'

                                               (1.1 * |mountain|
                                                < |water|)

Fig. 3 Qualitative representation of an image using relative size and Allen's calculus

3.1 Refinements of the qualitative representations

Several variants of the above qualitative representations were also investigated. One such case was a coarser chord representation with just three image areas namely, Top (T: top 3 rows in grid based image regions), Middle (M : for rows 4–7 of the image grid) and Bottom (B : constituted by rows 8–10); we also used an even coarser chord representation with just two chords: the top five and bottom five rows.

Since many Allen relations have zero count in a particular image, we also investigated a more coarse grained Allen-like representation where groups of Allen relations were collapsed to a set of five jointly exhaustive and pairwise disjoint relations (plus the "No" relation as before):

- BM combining "Before" and "Meet"
- Ov "Overlap" relation
- LG combining "Starts", "During", "Finishes", "Started by", "Contains", "Finished by" and "Equal"
- Ob "Overlapped by" relation
- AM combining "After" and "Met by"
- No if no Allen relation exists between two attributes of an image.

Top Region {
Region Middle {
Bottom Region {

sky fo liage
Sky fo liage
Sky fo liage
Sky fo liage
Sky fo liage
Sky fo liage
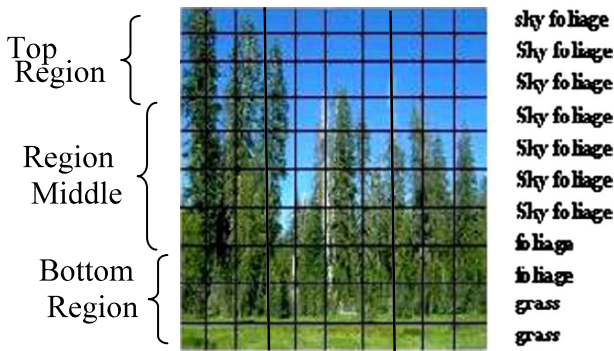Sky fo liage
fo liage
fo liage
grass
grass

Fig. 4 Qualitative representation of an image using the chord representation

4 IR based on qualitative similarity

We envisage an image retrieval system in which a query is specified either by giving an example image or by a symbolic query expressed in terms of the qualitative relations defined in the previous section. In the latter case the description is likely to be partial and a set of images will match the query, e.g. "retrieve images with rocks *meeting* water and water relatively *greater than* foliage". In the former case, we can compute a qualitative description of the image using one more of our qualitative schemes, but in this case it is more likely that no image will exactly match—this could also happen in the latter case. It would clearly be convenient to be able to retrieve images, which nearly match the query (which ever way it is specified). The problem is to define what "nearly matches" means, since in a qualitative representation we do not have raw numbers available. In the remainder of this section we define notions of qualitative similarity measure (QSM) for each the qualitative representations. The conceptual neighbourhood diagrams of each of the representations are used to calculate respective similarity measures.

- QSM for AllenRep: The conceptual neighbourhood diagram of Allen relations is presented in Fig. 5 above. Since the links in CND connect neighbouring relations—ones which are most similar, as one traverses links from a particular relation, the relations become progressively less similar. Thus if image 1 has sky < grass, and so does image 2, then they are identical (in this comparison), if image 3 has sky 'm' grass, then image 3 is similar to image 1, whilst if image 4 has sky 'o' grass, then image 4 is also similar to image 1 but not as similar as image 3, and so forth. Since there are 66 Allen relations in our description of an image, we have to find a way to combine the similarities of each pairwise comparison. The conceptual neighbourhood diagram for the Allen relations is already a partial order, and it is clear that the 66-cross product is much more partial in this respect. It is clearly desirable to have a total ordering. In order to achieve this we assign a weight of one to each arc in the conceptual neighbourhood diagram, and sum the number of arcs traversed across all 66 relations in order to transform one description into another (using the shortest route). Clearly we could assign non uniform weights to the different arcs but in the absence of any particular reason to do this a uniform weighting appears to be an obvious choice.

The situation where one of the relations from a particular attribute in a pair of images is "No" whilst the other is not, deserves some discussion—what should be the weight in this case (since "No" does not appear in the conceptual neighbourhood)? For a "No" relation, one or both of the concepts is not present in the image. One possibility is to choose a weight of seven (one more than the maximum weight otherwise in the Allen conceptual neighbourhood), though other choices could clearly also be used, and indeed we also experiment with the choice of zero[5]. This has further been verified by experimenting with values greater than one and between range 7–10; and choice of using a default weight of one came up with relatively better retrieval results. We envisage that in an implementation for an end user, this would be a parameter (perhaps a slider in the interface) so that the user can

---

[5] This was particularly motivated by image classes such as "lwm" where the set of concepts present in this class can vary considerably, and penalizing images with a different set of concepts to the query image had a great affect on the results for this category of images. A penalty weight of zero implies that the similarity of images is determined only by the relationship between common semantic concepts in the query and database images, and missing concepts do not contribute toward total penalty weight.
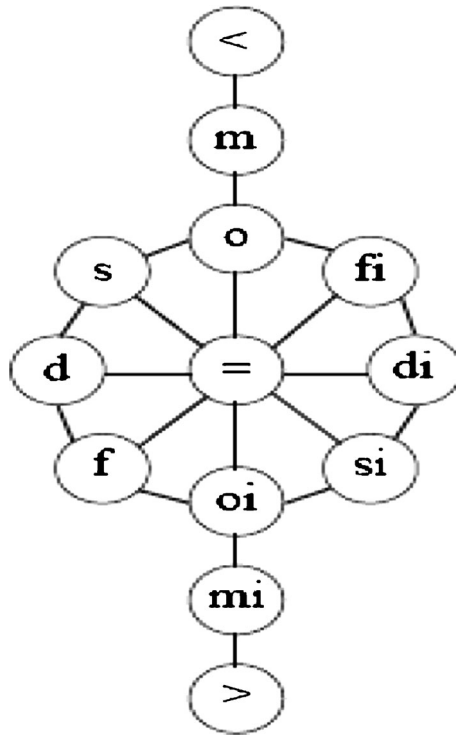
Fig. 5 Conceptual neighbourhood diagram for the interval calculus [7]

determine the effect of concepts not present in both images (though the current system takes into account only the default weights).

- QSM for RSizeRep: Turning to the relative size representation, the conceptual neighbourhood is much simpler with three nodes, one for each of the three relations, with $\approx$ neighbouring each of > and <. The maximum weight is two. For missing patch types we do not use a 'No' relation in this representation but rather use >, < if just one patch type is missing and $\approx$ if neither are present.

QSM for ChordRep: For the case of the chord representation, we can think of the conceptual neighbourhood as being equivalent to a complete lattice generated by the power set of patch types; effectively this means that the similarity is directly proportional to the number of insertions and deletions required to transform one chord into another.

- QSM for TouchRep: For the representation of spatial touching, there are just two nodes in the conceptual neighbourhood diagram (touching and not-touching) and a single link connecting them. We experimented with this representation, but converged on a similarity measure which also takes account of the degree of touching (i.e. taking a hybrid quantitative-qualitative approach). Each patch can touch up to eight other patches. For a pair of given patch types p1 and p2, we compute how many patches of type p1 touch a patch of type p2, and vice-versa for p2 and p1; the maximum of these two values is then recorded as one of the 66 attributes in this representation of an image. To compute the degree of similarity between two images using this representation, we simply take the sum of the absolute differences in each of the corresponding 66 values for each image. This

representation thus combines a very qualitative representation, touching, which is a purely topological relationship, with a metric measurement of its applicability to a particular image. Thus for example images with an extended sky-grass spatial connection will be more similar than ones where there is only a small amount of spatial connection between the two.

Following figure (Fig. 6) presents steps involved in image retrieval activity through a diagrammatic representation for ease of understanding.

Thus given a representation "R" with attributes $A_1{}^R, A_2{}^R, \ldots A_{|R|}{}^R$ and a function f^R (u,v) which gives the similarity between two attribute values u and v then the overall similarity S^R (x,y) between two images 'x' and 'y' in representation 'R' is given by:

$$S^R(x,y) = \sum_{i=1}^{|R|} f^R\left(A_i^R(x), A_i^R(y)\right) \qquad (1)$$

We can compute rank of an image 'y' in the database for query image 'x' as:

$$Rank^R(x,y) = z : S^R(x,z) < S^R(x,y) \qquad (2)$$

5 Results and evaluation

In this section, results for IR using the different representations described above are presented and evaluated. We have conducted experiments with each of the representations above individually and also in various combinations. To illustrate the results obtained, we first present (in Fig. 5) a sample query image and the top 5 results according to the qualitative similarity measures described in previous section—see Figs. 7, 8, 9, 10 and 11.
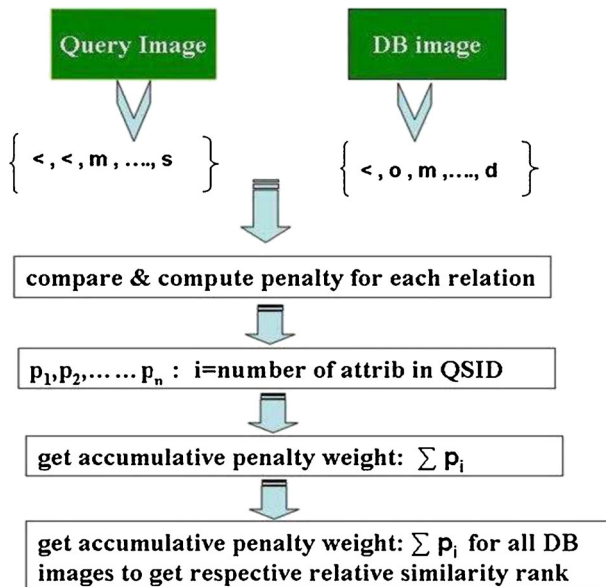

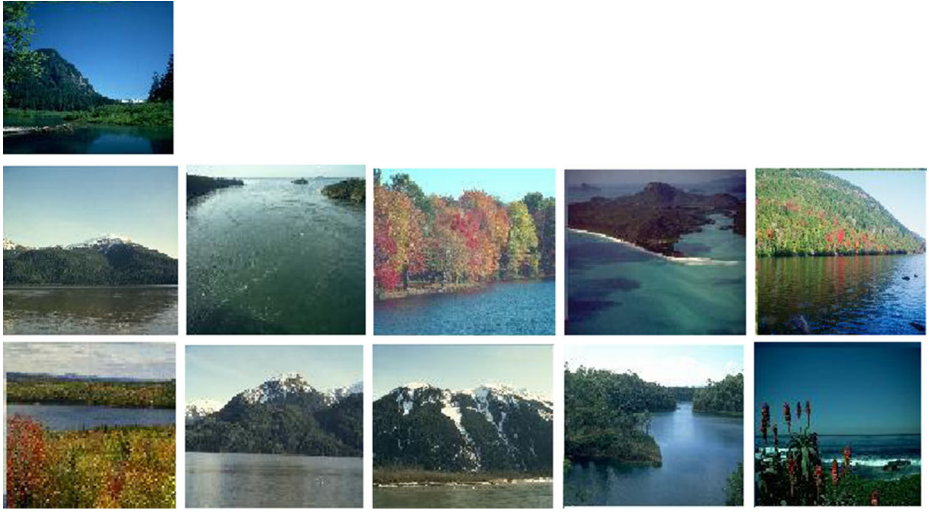Fig. 6 Tasks involved in image retrieval activity

Query Image

Fig. 7 Example image and top 10 images retrieved using Allen's grouped representation

The above querying does provide a visual appreciation of retrieval process but does not give any quantitative evaluation of the quality of the retrieval and we next turn to this question. To provide a more thorough quantitative analysis of the performance of the various representations, we used the following experimental setup. Each of the 700 images in the database was used as a query image in turn, and a similarity ordering computed for all the other 699 images. However this does not tell us whether images high in the ordering really are intuitively similar to the query image. As a proxy for an extensive user evaluation of each of these rank orderings, we use
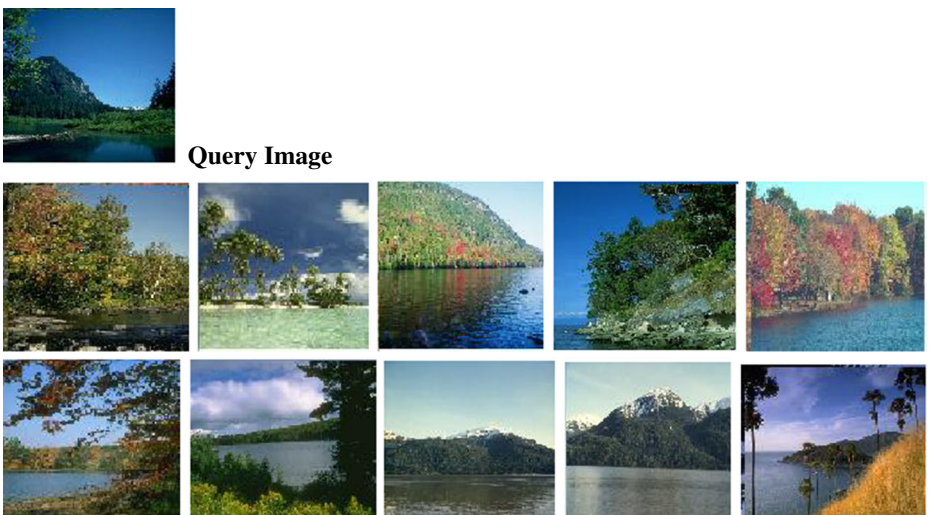

Query Image

Fig. 8 Example image and top 10 retrieved images using the chord representation

Fig. 9 Example image and top 10 images retrieved using Allen's representation

the hand assigned category labels used above and in [19] for supervised learning of category descriptions.

Given a query image in category c, we can evaluate the number and hence the percentage of images in the same category in the top k images in the rank ordering. For cases where the number of images of a particular category in the data base is less than k clearly 100 % scores cannot be achieved.

The number k may be user defined, or be determined by conditions such as how many images of a certain size fit on a user's screen, or could be determined by analysis of the
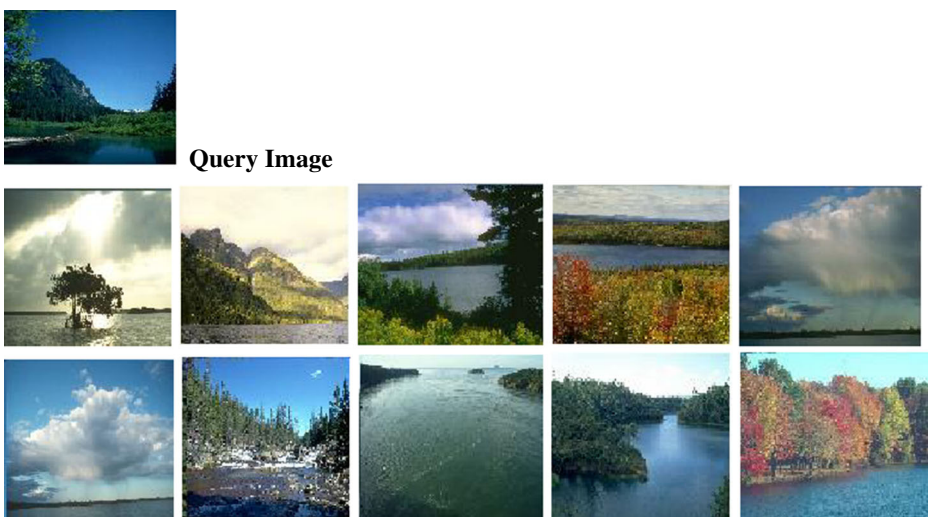

Fig. 10 Example image and top 10 images retrieved using the relative size representation

Query Image

Fig. 11 Example image and top 10 images retrieved using Allen's representation (with Zero Penalty Weight for "No" relation)

actual similarity values. In the figures below (Fig. 12, 13, 14 and 15), we histogram the similarity values for particular queries using various representations. It can be seen that typically there are "qualitative jumps" in the similarity values. These could be used to delineate the ordered list of images into "most similar", "fairly similar" and "least similar" sets. We have not experimented further with this approach in this paper or evaluated its cognitive plausibility (The legends 'Actual' represents frequency count of images while 'CF' represents the cumulative frequency of images in each bin—number of bins are plotted along X-axis while Y-axis represents number of images in each bin).

Table 1 gives a complete view, for each class, of the number of retrieved images of that class in the top ranked 20, 50 and 100 images, each row giving the values for a different representation. To show the effect of combining pairs of representations, we also show some hybrid representations, such as Allen_Size. The final row in each of the following three
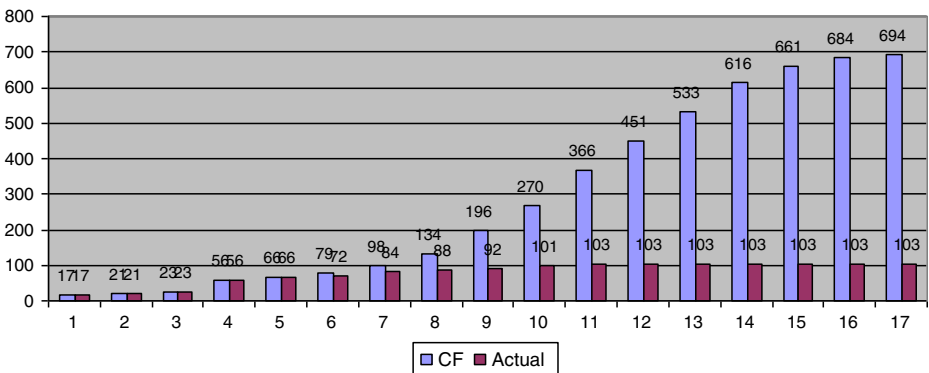

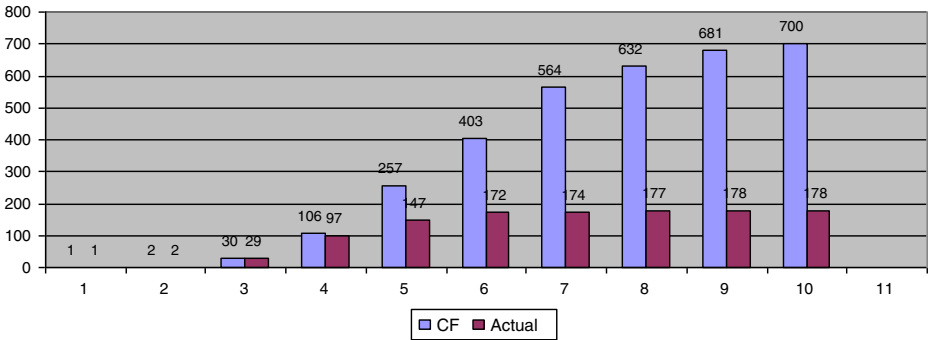**Fig. 12** Histogram for Allen representation

Fig. 13 Histogram for chord representation

tables shows the statistics when using the percentage of each semantic attribute as the image representation, as described Table 2 gives the same results but this time expressed as a percentage of the retrieved set. Clearly the number of actual occurrences of each image class in the database will affect the a priori probabilities of retrieving a particular image class; Table 3 thus presents the percentages of retrieved results relative to the actual number of each category of images in the database. In particular there are only 34 instances of Sky_Clouds in the database, so it is impossible to retrieve more than this number (in fact more than 34, assuming that one is used as the query image).

These results reveal the following interesting conclusions:

1. The recall rate clearly validates the measures of similarity used, since as the number of images retrieved increases, the accuracy of retrieved images goes down (measured by successive retrieved images of the same category).

2. The case of assigning zero weight to the "No" relation in image descriptions results in a decrease in recall value particularly in images corresponding to top most values in the ordered list of measure. This situation is improved if we
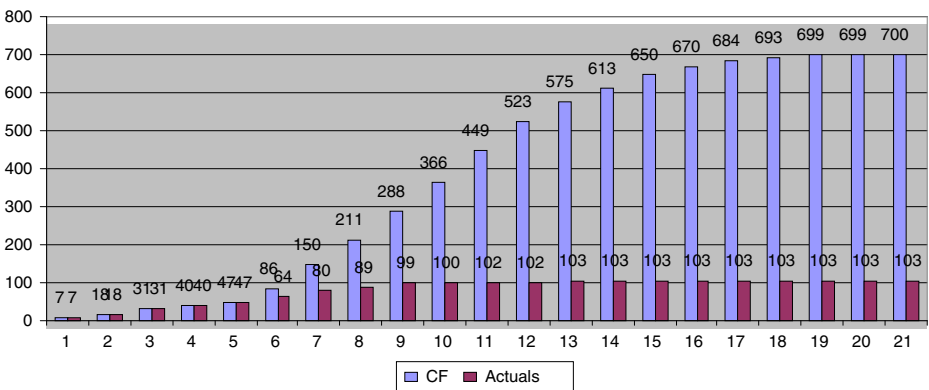

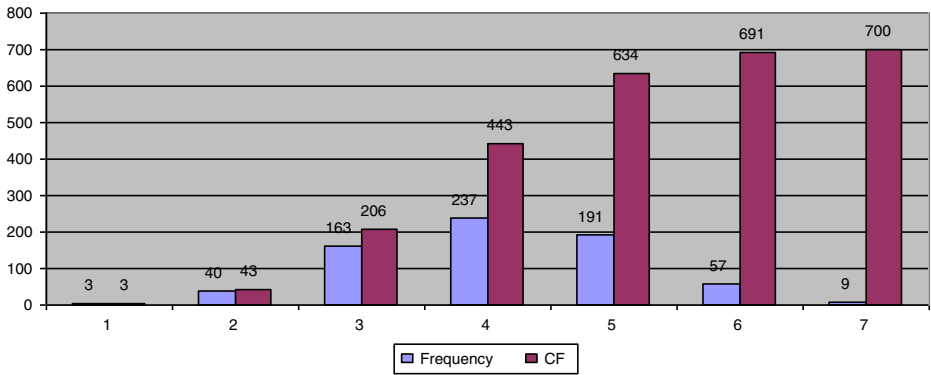**Fig. 14** Histogram for spatial touching representation

Fig. 15 Histogram for relative size representation

simultaneously consider relative size relation as well along with the Allen or Allen_LGp representations.

3. The chord representation performs relatively better than the other ones; it seems to be particularly suitable for measuring and comparing images. Arguable this is because it closely resembles the human cognition of similarity because human may describe or compare an image in terms such as an "image having sky in the top, foliage and water in the middle, water and sand at the bottom of image"—remembering that the semantic categories were assigned by a human (though without being aware of the possibility of subsequently using the chord representation (or indeed any of the others).

Table 1 Number of Images of each class retrieved in top 20, 50 and 100 retrieved images for all experiments

Categories/ Representations	Coasts: Out of			Field: Out of			Forest: Out of			LWM Out of			Sky_Clouds Out of			Waterscapes Out of		
	100	50	20	100	50	20	100	50	20	100	50	20	100	50	20	100	50	20
Allen only	53	28	14	29	21	13	85	50	20	74	41	20	15	3	3	46	26	13
Allen_LGp	47	27	12	27	17	12	80	50	20	64	42	20	8	3	3	50	28	15
Allen_Zero	33	26	16	29	22	18	55	37	20	36	14	8	8	4	3	15	4	3
Chord	17	10	6	49	39	20	81	50	20	91	47	19	33	33	20	59	31	13
Chord on TMB	16	5	3	43	37	18	68	49	20	47	38	14	33	31	20	47	25	14
Allen_Size	50	27	14	34	22	14	84	50	20	83	46	20	17	7	3	43	25	13
Allen_LGp_Size	49	30	12	30	21	13	84	50	20	86	44	20	15	4	3	46	26	13
Allen_Zero_Size	50	24	14	43	27	17	84	50	20	80	39	18	22	18	5	28	17	10
Chord_Size	44	26	14	42	29	17	89	50	20	88	49	20	26	17	12	39	22	12
Size only	48	28	13	38	26	16	86	50	0	86	49	20	25	16	10	36	21	10
Touch	36	19	11	29	15	7	68	50	20	88	48	20	23	17	14	20	14	10
Touch + Size	45	25	11	32	15	9	77	50	20	88	49	20	25	19	14	27	15	12
%ages	20	11	7	46	38	20	83	49	20	86	44	17	34	34	20	48	33	14

The Labels 'Allenonly' represent Allen Representation, 'Allen_LGp' represents Allen's variant described in Section 3.1, 'Allen_Zero' represents variant of Allen with zero penalty weight, 'Allen_Size' and others are hybrid representations by combining Allens and Size representations

Table 2 Recall percentages in top 20 and 100 images retrieved for all of the experiments

Representations	Coasts		Field		Forest		LWM		Sky_Clouds		Waterscapes		Over All
	Out of 20	Out of 100	Out of 20	Out of 100	Out of 20	Out of 100	Out of 20	Out of 100	Out of 20	Out of 100	Out of 20	Out of 100	
Allen only	70	53	60	29	100	85	100	74	15	15	65	46	50 %
Allen_LGp	60	47	60	27	100	80	100	64	15	8	75	50	46 %
Allen_Zero	80	33	90	29	100	55	40	36	15	8	15	15	29 %
Chord	30	17	100	49	100	81	95	91	100	33	65	59	55 %
Chord on TMB	15	16	90	43	100	68	70	47	100	33	70	47	42 %
Allen_Size	70	50	90	34	100	84	100	83	15	17	65	43	52 %
Allen_LGp_Size	60	49	70	30	100	84	100	86	15	15	65	46	52 %
Allen_Zero_Size	70	50	65	43	100	84	90	80	25	22	50	28	51 %
Chord_Size	70	44	85	42	100	89	100	88	60	26	60	39	55 %
Size only	65	48	80	38	100	86	100	86	50	25	50	36	53 %
Touch	55	36	35	29	100	68	100	88	70	23	50	20	44 %
Touch + Size	55	45	45	32	100	77	100	88	70	25	60	27	49 %
%ages	35	20	100	46	100	83	85	86	100	34	70	48	53 %

Table 3 Percentages of retrieved results relative to the total number of images in each category of database

Representations	Coasts		Field		Forest		LWM		Sky_Clouds		Waterscapes		Over all
	Out of 143	%age	Out of 128	%age	Out of 103	%age	Out of 178	%age	Out of 34	%age	Out of 114	%age	%age
Allen only	69	48	37	29	86	84	108	61	3	9	48	42	46
Allen_LGp	66	46	33	26	80	78	105	59	3	9	56	49	45
Allen_Zero	42	29	35	27	57	55	54	30	3	9	17	15	28
Chord	22	15	57	45	82	80	127	71	30	88	63	55	59
Chord on TMB	23	16	52	41	70	68	80	45	24	71	52	46	48
Allen_Size	62	43	42	33	86	83	110	62	3	9	49	43	46
Allen_LGp_Size	61	43	40	31	86	83	109	61	3	9	50	44	45
Allen_Zero_Size	61	43	45	35	85	83	110	62	17	50	29	25	50
Chord_Size	59	41	50	39	90	87	124	70	16	47	44	39	54
Size only	60	42	48	38	87	84	123	69	15	44	36	32	52
Touch	50	35	35	27	68	66	114	64	14	41	23	20	42
Touch + Size	60	42	41	32	78	76	121	68	16	47	29	25	48
%ages	26	18	54	42	84	82	135	76	31	91	51	45	59

Table 4 Recall percentages in top 5 and 10 pooled images from each of 4 representations

Class of Q_Image		Coasts	Fields	Forest	LWM	Sky_clouds	Waterscapes
Recall values	Top 5	69 %	83 %	100 %	100 %	46 %	50 %
	Top 10	54 %	78 %	100 %	100 %	50 %	52 %

4. The representation 'relative size' performs surprisingly well, given the low information content.
5. The collapsed version of the Allen representation, namely Allen_LGp, is not as good as the best representations. The obvious conclusion to draw is that the representation is too coarse to successfully distinguish somewhat dissimilar images.
6. The touch based representation does not perform particularly well either—again we hypothesise that it does not encode sufficient information to be able to adequately distinguish cognitive similarity in the image dataset.

As can be seen above in Figs. 12, 13, 14 and 15, we have considered combining the representations (e.g. Allen and relative size). We now consider hybrid representations further. In the binary combinations above, the two representations carried equal weight (although we considered weighting them corresponding to their overall performance). Since different representations perform better in different categories (and bearing in mind that we assume we do not know the category of an image—we are using this information here purely for evaluation purposes), we experimented with combinations of four different qualitative representations.

5.1 IR using voting based QSRs

We experimented with a number of voting schemes to aid and improve the retrieval process using multiple representations[6]. The selected representations were Allen, relative size, chord and touch. There has been number of approaches in image categorization research involving bagging/boosting while in image retrieval, multiple query processing or use of low level and semantic labels has been used to improve the retrieval accuracy. We have investigated following four voting approaches based on combining the respective penalty weights of images in individual representations[7], and on combining the ranks of retrieved images in each selected qualitative representation:

V_1: Compute…

$$S^{V_1}(x,y) = \sum_{R=1}^{r=4} S^r(x,y) \tag{3}$$

for each image in the DB for a query 'x' and then sort in ascending order.

V_2: Compute…

$$S^{V_2}(x,y) = \Phi_{r=1}^{r=4} S^r(x,y) \tag{4}$$

for each image in the DB for a query 'x' and then sort in ascending order:(a variant of V1 and Φ is "Min" function). Although the weights within in each representation may be

---

[6] Of course each representation can itself be a hybrid representation with the 66 attributes (or whatever number of attributes used in the particular representation) combining together to assign an overall similarity to an image pair.
[7] This is thus the same as the combination used within each representation, as mentioned in the previous footnote.

Table 5 Recall percentages in top 20 and total number of images in each category of database for AllRep RowSum Weighted Voting

Class of query image		Coasts 143	Fields 128	Forest 103	Lwm 178	Sky_clouds 34	Waterscapes 114
Recall values	top 20	70 %	65 %	100 %	100 %	20 %	65 %
	Top k	46 %	29 %	85 %	61 %	28 %	43 %

regarded as comparable, it is arguable as to whether this also holds with respect to the weights in other representations. We thus investigated schemes based solely on the rank within each of the four representations.

V_3: Compute…

$$S^{V_3}(x,y) = \sum_{r=1}^{r=4} rank^r(x,y) \qquad (5)$$

for each image in the DB for a query 'x' and then sort in ascending order.

V_4: Compute…

$$S^{V_4}(x,y) = \Omega_{r=1}^{r=4} rank^r(x,y) + \Psi_{r=1}^{r=4} rank^r(x,y) \qquad (6)$$

where "Ω" and "Ψ" compute the maximum and 2nd highest values respectively.

In the first case of simple voting, the top 5 and 10 retrieved images from each of the four representations with respect to a query image were pooled together, removing duplicates. For the evaluation we then counted how many images in the resulting pool were in the same category as the query image. The results of this approach are given in Table 4.

In order to count for the accumulative effect of penalty weights in all of the 4 representations mentioned in above paragraph and also the overall ranking of an image in the list of database images, several other kinds of weighted voting schemes were investigated (V_1–V_4):

In voting scheme V_1, the individual weights of each image in the data base in each of the four representations were computed, and the all four weights summed. The database images were then sorted in ascending order of this "rowsum" weight for each representation for each. See Table 5.

A variant of the rowsum scheme is to take the minimum weight (MinWt) of each image across all of the four representations (Voting scheme V_2). Thus an image which has a low rank in all four representations will still have a low rank, but an image which is selected out as very similar buy just one representation will be be penalised by representations which weight it as rather un-similar to the query image. The results for this approach are presented in Table 6.

Although the weights within in each representation may be regarded as comparable, it is arguable as to whether they should be compared with the weights in other representations. We thus investigated schemes based solely on the rank within each of the four representations (V_3 and V_4). In V_3, ranks of each image in each of the sorted lists of the four individual representations were added together to have the "RankSum" of each image—e.g. an image ranked 2,5,15 and 40 in the four representations would receive total "vote" of 62. The retrieved

Table 6 Recall percentages in top 20 and total number of images in each category of database for AllRep RowSum Weighted Voting using MinWt

Class of query image		Coasts 143	Fields 128	Forest 103	Lwm 178	Sky_clouds 34	Waterscapes 114
Recall values	top 20	55 %	100 %	100 %	100 %	90 %	65 %
	Top k	19 %	45 %	82 %	73 %	82 %	54 %

Table 7 Recall percentages in top 20 and total number of images in each category of database for RankSum Weighted Voting

Class of query image		Coasts 143	Fields 128	Forest 103	Lwm 178	Sky_clouds 34	Waterscapes 114
Recall values	top 20	60 %	45 %	100 %	100 %	70 %	60 %
	Top k	41 %	39 %	89 %	73 %	44 %	41 %

images were then sorted in ascending order corresponding to a decreasing relative similarity of database images to the query image. The results of this approach are presented in Table 7.

Finally, similarly to voting scheme MinWt, rather than combine the ranks from all four representations, the overall weight for each image was computed as the sum of its two best weights across the four representations (V_4). The results of this voting scheme is displayed in Table 8.

The results above suggest the following conclusions:

- The purely qualitative approaches perform as well as or even slightly better in some cases than the quantitative ones. The qualitative approaches, though, have added advantage that these approaches also allow retrieval based on simple linguistic descriptions using qualitative descriptions over the semantic attributes.
- The voting schemes based on accumulative weighted votes and ranksum weighted votes perform better than the individual approaches.
- The overall accuracy of the retrieval process compared with the actual class labels is not entirely a fair evaluation due to the fact that most of the images may be categorized as either "lwm" or "coast"—i.e. most of the images in the data base have some aspects of "lwm" or "coast", and arguably it is a matter of degree or personal preference when an lwm with sky above becomes a "sky_clouds" for example. Similarly, there is lot of confusion in images categorised in classes like "fields" and "sky_clouds". This fact was also established in [19, 31] while learning the class descriptions using the same image descriptions.
- The minimum weights (MinWt) approach in the AllRep RowSum Weighted Voting scheme performs much better in the top 20 and the top k (k is equal to total number of images in each category of mages in database) as it is based on minimum row weight of an image out of penalty weights corresponding to four representations chosen. This approach filters out relatively 'more' similar image to a query image from list of retrieved images of all representations. The second best voting scheme, namely RankSum Weighted Voting, performs slightly better in the case of using best two ranks out of four (Table 8 rather than Table 7), though it is much better in the recall of the top 20 rather than the top k.
- It can be seen that coasts and waterscapes do relatively badly compared to the other categories in many of the representations, which is not altogether surprising from a semantic/intuitive viewpoint. If these two categories are combined into a single category

Table 8 Recall percentages in top 20 and total number of images in each category of database for RankSum Weighted Voting

Class of Query Image		Coasts 143	Fields 128	Forest 103	Lwm 178	Sky_clouds 34	Waterscapes 114
Recall values	top 20	75 %	75 %	100 %	100 %	55 %	65 %
	Top k	41 %	36 %	87 %	72 %	47 %	42 %

then the rate of accuracy improves. For example the figures for the top 20 and top 100 images for this combined category are as follows for a selection of representations:

Allen:	20/20	87/100
Relative Size:	17/20	72/100
Chord:	20/20	87/100
T_n_T (Touching):	10/20	29/100
RowSum with MinWt:	20/20	87/100
Rank_Sum (Top two ranks out of 4):	16/20	75/100

6 Conclusions and further work

We have presented an approach to image retrieval based on semantic knowledge and qualitative spatial image descriptions. The approach does not rely either on segmentation techniques applied directly or on low level image features for an image description. We have presented similarity measures of the qualitative spaces based on the conceptual neighbourhoods that typically accompany qualitative calculi. We have presented the results for a variety of qualitative description languages and several combinations of these. We are not necessarily arguing that these are the best languages either for this particular data set or in general. It is the overall approach we present which we believe is the most important result of our research. We have also presented a variety of voting schemes for combining representations and evaluated their success on the image dataset.

The evaluation was based on a hand labeled categorization which although it has some disadvantages does provide a cognitive basis for evaluating the retrieval results.

We have also suggested the use of histogram analysis to categorize retrieved results into categories of similarity.

A variety of further work suggests itself including the evaluation on other data sets, using actual user analysis to evaluate the results, experimentation with other qualitative calculi, and combining qualitative and quantitative representations. We already have a prototype user interface to an image retrieval system based on the ideas presented here; this could be further improved to provide a flexible interface based on query by image or by qualitative description, or a combination of the two, with the user free to select the kinds of descriptions, similarity measures and voting schemes most appropriate to their needs. The analysis here provides the basis for reasonable default choices.

References

1. Aghrabi Z, Makinouchi A (2003) Semantic approach to image database classification and retrieval. NII journal, No. 7 (2003.9)
2. Allen JF (1983) Maintaining knowledge about temporal intervals. In Communications of ACM, 26(11)
3. Bonan L, Fonsesca F (2006) TDD: A comprehensive model for qualitative spatial similarity assessment. J Spatial Cog Comput 6(1):31–62

4. Bradshaw B (2000) Semantic based image retrieval: A probabilistic approach. ACM Multimedia
5. Burns HT, Egenhofer MJ (1996) Similarity of spatial scenes. In: Kraak J.-M, and Molenaar M (eds) Seventh international symposium on spatial data handling, Delft, The Netherlands, Taylor & Francis, London, pp. 173–184
6. Ciocca G, Cusano C, Santini S, Schettini R (2011) Halfway through the semantic Gap: prosemantic features for image retrieval. Infor Sci, Elsevier 181:4943–4958
7. Cohn AG, Hazarika SM (2001) Qualitative spatial representation and reasoning: an overview. Fundamenta Informaticae 46(1–2):1–29
8. Deb S, Zhang Y (2004) An overview of content-based image retrieval techniques. In Proceedings of 18th International Conference on Advanced Information Networking and Application (AINA'04), Vol.1, pp. 59–64
9. Depalov D, Pappas T, Li D, Gandhi B (2006) Perceptually based techniques for semantic image classification and retrieval. In: Rogowitz BE, Pappas TN, Daly SJ (eds) Human vision and electronic imaging, XI. Proceedings SPIE, Vol. 6057, (San Jose, CA)
10. Enser P, Sandom C (2003) Towards a comprehensive survey of the semantic gap in visual image retrieval. In International Conference on Image and Video Retrieval, Springer Lecture Notes in Computer Science 27–28: 279–287
11. Freksa C (1992) Temporal reasoning based on semi-intervals. Artif Intell 54(1–2):199–227
12. Howe N (2003) A closer look at boosted image retrieval. In International Conference on Image & Video Retrieval, Vol. 2728, LNCS, pp. 61–70
13. Hyvonen E, Styrman A, Saarela S (2002) Ontology-based image retrieval. In HIIT publications, No. 2002–03, pp. 15–27, Helsinki Institute of Information Technology, Helsinki, Finland
14. Iqbal Q, Aggarwal JK (2002) CIRES: A system for content-based image retrieval in digital image libraries. Invited Session on CBIR: Techniques and applications, international conference on control, automation, robotics and vision (ICARCV), pp. 205–210
15. Madugunki M, Bormane DS, Bhadoria S, Dethe CG (2011) Comparison of different CBIR techniques. In Proceedings of 3rd International Conference on Electronics Computer Technology (ICECT), 4: 372–375
16. Maron O, Lakshmi Ratan A (1998) Multiple-instance learning for natural scene classification. In Procdings of 15th International Conference on Machine Learning, PP 341–349
17. Mörchen F, Ultsch A (2005) Discovering temporal knowledge in multivariate time series. In: Weihs C, Gaul W (Eds) In Classification; The Ubiquitous Challenge, Proceedings 28th Annual Conference of the German Classification Society (GfKl 2004), Dortmund, Germany, Springer, Heidelberg, pp. 272–279
18. Picard R, Minka T (1995) Vision texture for annotation. ACM J Multi Syst
19. Qayyum ZU, Cohn AG (2006) Qualitative approaches to semantic scene modelling and retrieval. In Research and Development in Intelligent Systems XXIII, Springer-Verlag
20. Qayyum ZU, Cohn AG (2007) Image retrieval through qualitative representations over semantic features. In Proceedings of 18th British Machine Vision Conference (BMVC), 1–10
21. Randell DA, Cui Z, Cohn AG (1992) A spatial logic based on regions and connection. In Proceedings of 3rd International Conference on Knowledge Representation and Reasoning, pages 165–176, Morgan Kaufmann
22. Rasiwasia N, Vasconcelos N, Moreno PJ (2006) "Query by Semantic Example". In: Sundaram H et al. (eds) CIVR 2006, LNCS 4071, pp. 51–60
23. Sebe N, Lew MS, Zhou X, Huang TS, Bakker EM (2003) The state of the art in image and video retrieval. In International Conference on Image & Video Retrieval, Vol. 2728, LNCS, pp. 1–8
24. Serrano N, Savakis A, Luo J (2004) Improved scene classification using efficient low level features and semantic cues. Patt Recog 37(9):1773–1784
25. Singha M, Hemachandran K (2012) Content based image retrievel using color and texture. Signal and Image Processing: An International Journal (SIPIJ), 3(1)
26. Singhai N, Shandilya S (2010) A survey on content based image retrieval systems. IJCA 4(2):22–26
27. Siqnig W, Gang X, Geng W, Gouping Z (2011) Digital-image retrieval based on shape-feature and MPEG-7 standard. In Proceedings of International Conference on Consumer Electronics, Communication and Networks, 3095–3097
28. Sun JY, Sun ZX, Zhou RH, Wang HF (2002) A semantic-based image retrieval system: VISENGINE. In Proceedings of First International Conference on Machine Learning and Cybernetics, Beijing
29. Town C, Sinclair D (2000) Content-based image retrieval using semantic visual categories. Tech. Report 2000.14, AT&T Laboratories Cambridge
30. Vailaya A, Figueiredo M, Jain A, Zhang HJ (2001) Image classification for content-based indexing. IEEE Trans Image Proc 10(1):117–130
31. Vogel J, Schiele B (2006) Semantic modelling of natural scenes for content-based image retrieval. Intern J Comp Vision, Springer Netherlands. doi:10.1007/s11263-006-8614-1
32. Wang W, Song Y, Zhang A (2002) Semantic retrieval by content and context of image regions. In Proceedings of 15th International Conference on Vision Interface (VI'02)

Prof. Qayyum is currently working at National University of Computer & Emerging Sciences, Islamabad. He has over 20 years of experience in teaching and research. His major areas of interest include Data Mining, Knowledge Representation and Information Retrieval.